Evaluating the Augmented Reality Human-Robot Collaboration System

Scott A. Green, J. Geoffrey Chase, XiaoQi Chen
Department of Mechanical Engineering
University of Canterbury, Christchurch, New Zealand
Email: scott.green, xiaoqi.chen, geoff.chase@canterbury.ac.nz

Mark Billinghurst
Human Interface Technology Laboratory, NZ (HITLab NZ)
University of Canterbury, Christchurch, New Zealand
Email: mark.billinghurst@canterbury.ac.nz

***Abstract*-This paper discusses an experimental comparison of three user interface techniques for interaction with a mobile robot located remotely from the user. A typical means of operating a robot in such a situation is to teleoperate the robot using visual cues from a camera that displays the robot's view of its work environment. However, the operator often has a difficult time maintaining awareness of the robot in its surroundings due to this single ego-centric view. Hence, a multi-modal system has been developed that allows the remote human operator to view the robot in its work environment through an Augmented Reality (AR) interface. The operator is able to use spoken dialog, reach into the 3D graphic representation of the work environment and discuss the intended actions of the robot to create a true collaboration. This study compares the typical ego-centric driven view to two versions of an AR interaction system for an experiment remotely operating a simulated mobile robot. One interface provides an immediate response from the remotely located robot. In contrast, the Augmented Reality Human-Robot Collaboration (AR-HRC) System interface enables the user to discuss and review a plan with the robot prior to execution. The AR-HRC interface was most effective, increasing accuracy by 30% with tighter variation, while reducing the number of close calls in operating the robot by factors of ~3x. It thus provides the means to maintain spatial awareness and give the users the feeling they were working in a true collaborative environment.**

I. Introduction

Interface design for Human-Robot Interaction (HRI) is becoming one of the toughest challenges that the field of robotics faces [1]. As HRI interfaces mature it will become more common for humans and robots to work together in a collaborative manner. With this idea in mind, a system has been developed that allows humans to communicate with robotic systems in a natural manner through spoken dialog and gesture interaction, the Augmented Reality Human-Robot Collaboration (AR-HRC) system [2].

Augmented Reality (AR) blends virtual 3D graphics with the real world in real time [3]. AR allows real time interaction with the 3D graphics, enabling the user to reach into the augmented world and manipulate the 3D objects directly as if they were real objects. The virtual graphics used in this work depict the robot in a common workspace that both the human and robot can reference. Providing the human with an exo-centric view of the of the robot and its surroundings enables the human to maintain situational awareness of the robot and gives the human-robot team the ability to ground their communication [4] and create a truer collaboration for complex tasks.

This paper clinically evaluates the AR-HRC system. The task was to guide a simulated mobile robot through a predefined maze. Three user interfaces were compared for performance and collaboration. One interface was a typical teleoperation mode with a single ego-centric camera feed from the robot. A second interface was a limited version of the AR-HRC system that allowed the user to see the robot in its work environment through the AR interface, but did not provide any means of pre-planning or review of the robot's intended actions. The third interface was the full AR-HRC system that allowed the user to view the robot in the AR environment and to use spoken dialog and gestures to work with the robot to create and review a plan prior to execution.

The dependent variables measured in the experiments were the time to completion, accuracy in reaching predefined points in the maze, the number of impending and actual collisions with objects. In addition, the dialog used throughout the experiment was analyzed. Subjective questionnaires were administered after each of the three trials along with a final questionnaire upon completion of the entire experiment comparing the three interfaces tested.

II. Related Work

Pioneering work from Milgram *et al* [5] highlighted the need for combining the attributes humans are good at with those that robots are good at to create an optimized human-robot team. For example, humans are good at deictic referencing, such as using 'here' and 'there', whereas robotic systems need highly accurate discrete positional information. Milgram *et al* pointed out the need for HRI systems to convert the methods considered natural for human communication to the precision required for machine information.

Bolt's work "Put-That-There" [6] showed that gestures combined with natural speech lead to a more natural human-machine interface. Skubic *et al.* [7] conducted a study on human-robotic interaction using a multimodal interface. The result was natural human-robot spatial dialog enabling the robot to communicate obstacle locations relative to itself and receive verbal commands to move to an object it had detected.

Collaborative control was developed by Fong *et al* [8] for mobile autonomous robots. The robots work autonomously until they run into a problem they can't solve. At this point, the robots ask the remote operator for assistance, allowing robot autonomy to vary as needed. Results showed that robot

performance increases with the addition of human skills, perception and cognition, and benefit from human advice and expertise

Bowen *et al* [9] and Maida *et al* [10] showed through user studies that the use of AR resulted in significant improvements in robotic control performance. Similarly, Drury *et al* [11] found that augmented real-time video with pre-loaded map terrain data resulted in a statistical improvement in comprehension of 3D spatial relationships over using 2D video alone for operators of Unmanned Aerial Vehicles (UAVs). The augmented video resulted in increased situational awareness of the activities of the UAV.

Finally, Augmented Reality (AR) can create a more ideal environment for human-robot collaboration [12]. In a study of the performance of human-robot interaction in urban search and rescue, Yanco *et al.* [13] identified the need for situational awareness of the robot and its surroundings. In particular, the AR-HRC system significantly benefits from the use of AR technology to convey visual cues that enhance communication and grounding, enabling the human to have a better understanding of what the robot is doing and its intentions.

The multimodal approach employed in developing the AR-HRC system in this work combines spatial dialog, gesture and a shared reference of the work environment. The shared visual reference is accomplished using AR. The human and robot are thus able to discuss a plan, review the plan and then once a plan has been agreed upon, send it off for execution.

III. Experimental Design

The task for the user study was to guide a simulated robot through a predefined maze. Three conditions were used:

- ***Immersive Test***: A typical teleoperation mode with a single ego-centric view from the robot's onboard camera.
- ***Speech and Gesture no Planning (SGnoP)*:** A limited version of the AR-HRC system that allowed the user to see the robot in its work environment in AR and interact with the it using speech and gesture, but without pre-planning and review of the robot's intended actions.
- ***Speech and Gesture with Planning, Review and Modification (SGwPRM):*** The full AR-HRC system that allowed the human to view the robot in the AR environment, use spoken dialog and gestures to work with the robot to create and review a plan prior to execution.

The three conditions are, therefore, distinguished by increasing levels of collaboration or communication channels.

Ten participants were run through the experiment, seven male and three female. Ages ranged from 28 to 80 and all participants were working professionals. Seven of the participants were engineers while the other three had non-scientific backgrounds. Overall, the users rated themselves as not familiar with robotic systems, speech systems or AR.

The first step of the experiment was to have each participant fill out a demographic questionnaire to evaluate their familiarity with AR, game playing experience, age, gender and educational experience. Since speech recognition was an integral part of the experiment it was necessary to have each participant run through a speech training exercise. This exercise created a profile for each user so that the system was better able to adapt to the speech of the individual participant.

The objective of each trial was then explained to the participants. They were told that they would be interacting with a mobile robot to get it through the predefined maze. The maze contained a defined path for the robot to follow and various obstacles, around which the robot would need to maneuver. The participants were told that the robot must arrive at each of the numbers on the map as this goal was going to be a measure of accuracy for the test. Other parameters measured were impending collisions, actual collisions and time to completion. These metrics thus cover performance, accuracy and cost in time, as the interface increases in collaborative capability and interaction.

It was explained to the participants that the robot was located remotely. Thus, when the robot was directly driven a time delay would be experienced. Therefore, any delay in reaction of the simulated robot was not the system failing, but was the result of the time taken for the commands to reach the robot and the update from the robot to arrive back to the user. This delay thus mimics the situation experienced in any teleoperation, particularly for space-based applications.

The experimental setup used was a typical video see through AR configuration. A webcam attached to an eMagin Z800 Head Mounted Display (HMD) [14] and the HMD were connected to a laptop PC running ARToolKit [15] based software. Vision techniques were use to identify unique markers in the user's view and align the 3D virtual images of the robot in its world to these markers. This augmented view was presented to the user in the HMD. Fig. 1 shows a participant using the AR-HRC system during the experiment.

Figure 1. A participant using the AR-HRC system. The image on the monitor is what is being displayed to the user in the HMD.

The same sequence of events took place for each trial. Before each trial the participant practiced using the system to become familiar with the interface for that particular condition. The user also practiced any speech specific to that trial. Once the user felt comfortable with the interface the trial was run.

When a trial was complete the user was given a subjective questionnaire to determine if they felt that they had a high level of spatial awareness during the trial. The user was also questioned about whether they felt present in the robot's world and their view of the robot as a partner. The participants were also asked to list what they liked and disliked about the condition. This questionnaire was exactly the same for all three trials.

At the end of the experiment, after the participant had completed all three trials, a subjective questionnaire was given so the user could compare the three conditions. The post trial questionnaires discussed previously referred only to the trial that had just been completed. The subjective questioning was conducted in this manner to let the user express their feeling about each condition individually and then compare the three conditions upon completion of the full experiment. The order of the conditions was counterbalanced between users to avoid sequencing affecting the experimental results [16].

The Immersive Test simulated the direct teleoperation of the robot with visual feedback to the user displaying the view that the robot saw through its camera. This view provided the user with an ego-centric view of the robot's environment. User interaction included keyed input for robot translation and rotation. The view the user experienced can be seen in Fig. 2.

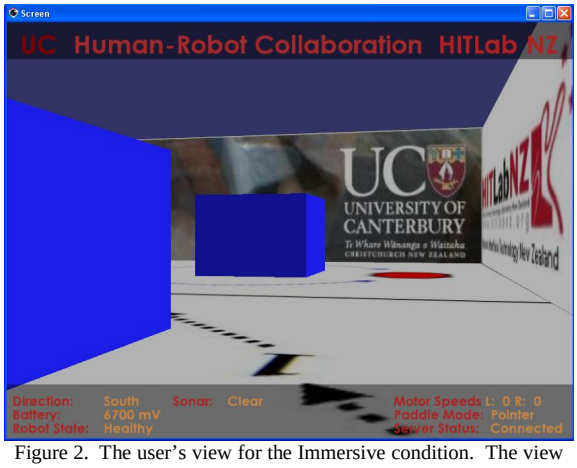


Figure 2. The user's view for the Immersive condition. The view shown is that from the robot.

The SGnoP condition provided the user with a 3D graphic of the robot and maze. The participant was able to use spatial dialog coupled with paddle gestures to interact with the graphical world of the robot in the AR environment. Using a handheld paddle, the participant was able to point to a 3D location on the maze and instruct the robot to "go there" or select an object and instruct the robot to "go to the right of that". The robot responded immediately to the verbal commands given after a time delay for the simulation of a remotely located robot. The speech was one-way in that the system in this condition understood the user's spatial dialog but did not respond verbally, thus offering input without collaboration. The view provided to the participant can be seen in Fig 3.

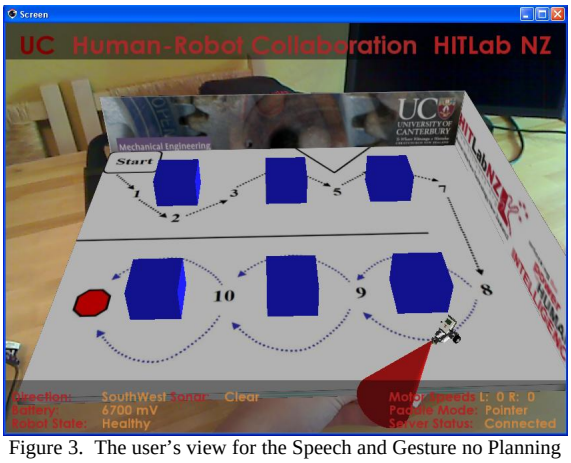


Figure 3. The user's view for the Speech and Gesture no Planning condition.

The user's view for the SGwPRM condition can be seen in Fig 4. This condition included all the features of the SGnoP condition but also allowed the participant to use spatial dialog to create a plan with the robot. The user was able to select a goal location and then assign way points for the robot to follow to arrive at the goal destination. The user could interactively modify the plan by adding or deleting way points. The plan was displayed to the user in the AR environment thus making it easy to determine if the intentions of the robot matched those of the user before any motion commands were executed by the robot. The robot participated in the dialog by responding to the user verbally for each interaction and alerting the user verbally when the robot came close enough to an object that the robot "thought" it would collide.

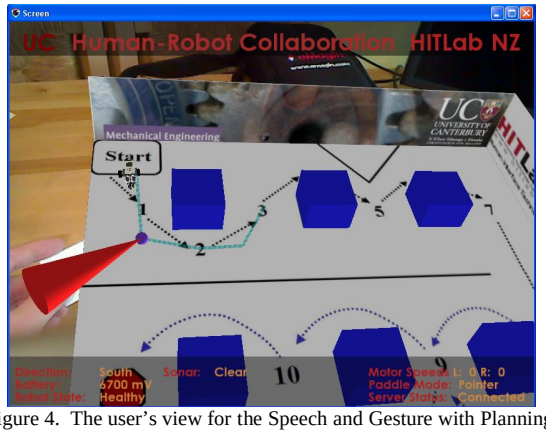


Figure 4. The user's view for the Speech and Gesture with Planning, Review and Modification condition. The user is creating a plan (blue line) that includes various waypoints through the use of spatial dialog and gesture.

IV. Results

The ten participants each performed three tasks, one for each condition. Each trial yielded a measure of time to completion, impending collisions, number of collisions and accuracy in reaching each of the ten defined locations on the map. An impending collision was defined as any time the robot came within a predefined threshold of an object. A warning was given to the user that an object was close enough to the robot

that a human perspective was needed to determine if the current course of action was clear.

There was a significant main effect of experiment condition on the average task completion times, see Fig. 5, with an ANOVA test finding ($F_{2,27}$ = 9.83, p< 0.05). Bonferroni correction [17] identifies which means are significantly different, and is used in this analysis when the ANOVA test shows a significant main effect of experiment condition. Pairwise comparison with Bonferroni correction (p < 0.05) revealed significant differences between the SGwPRM and the other two conditions. However, there was no significant difference between SGnoP and the Immersive conditions. Users in the Immersive condition performed faster than the other two conditions with a mean completion time of 331.60 seconds (se = 36.72).

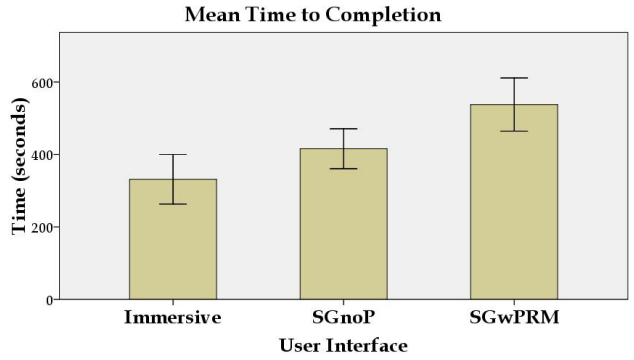
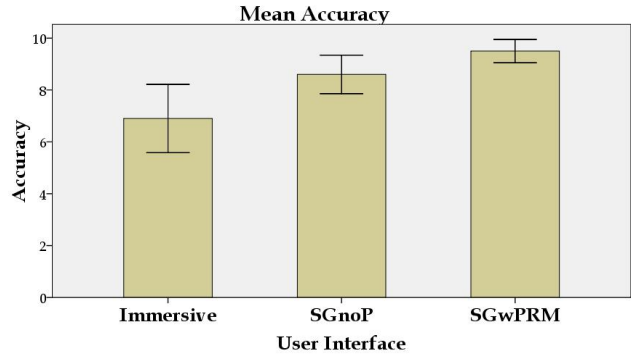


Figure 5. Mean time to completion

The experiment condition also significantly affected accuracy, see Fig. 6, with an ANOVA test finding ($F_{2,27}$ = 8.44, p< 0.05). Pairwise comparison with Bonferroni correction (p < 0.05) revealed significant differences between the SGwPRM and Immersive conditions but no significant differences between the SGnoP and the other two conditions. The SGwPRM performed the best by arriving at an average of 9.50 out of 10 defined locations (se = 0.22).



Number of goal locations reached out of 10.

Figure 6. Mean accuracy.

There was a significant main effect of experiment condition on the average number of close calls, see Fig. 7, with an ANOVA result of ($F_{2,27}$ = 13.10, p< 0.05), but no significant effect on the number of collisions. Pairwise comparison using Bonferroni correction (p < 0.05) showed significant differences for close calls between the Immersive condition and the other two conditions. There was no significant difference between SGnoP and SGwPRM. The SGwPRM condition performed best with a mean number of close calls of 3.60 (se = 1.01).

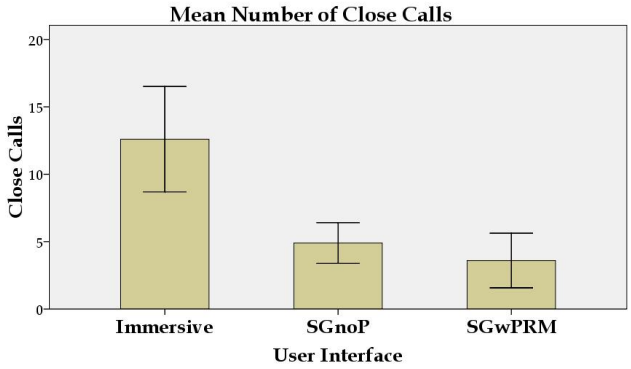


Figure 7. Mean number of close calls.

The answer for each post trial question was given on a Likert scale of 1-7 (1 = disagree completely, 7 = agree completely) and analyzed using an ANOVA test. If necessary, post-hoc analysis was performed using Bonferroni correction (p < 0.05). The results of the questionnaires for the individual trials (PT) are presented first.

- PTQ1: *I knew exactly where the robot was at all times.* There was a significant difference between conditions ($F_{2,27}$ = 7.43, p < 0.05). Pairwise comparison showed a significant effect between the Immersive condition and the other two conditions, but no significant effect between the SGnoP and SGwPRM conditions. Users felt that they maintained situational awareness best using the SGwPRM condition.
- PTQ2: *The interface was intuitive to use.* There was no significant difference between the conditions.
- PTQ3: *The robot was a member of my team as we completed the task.* There was a significant difference between the conditions ($F_{2,27}$ = 6.07, p < 0.05). Pairwise comparison revealed a significant effect between the Immersive condition and the two others. There was no significant difference between the SGnoP and SGwPRM conditions. The users felt that the robot was a member of their team in the SGwPRM condition.
- PTQ4: *I felt a sense of being present in the robot's world*. There was no significant difference between the conditions.
- PTQ5: *I was always aware of how close the robot was to objects in its environment.* There was no significant different between the three conditions.
- PTQ6: *I felt like the robot was just a tool and not a collaborative partner.* There was a significant difference between conditions ($F_{2,27}$ = 5.68, p < 0.05). Pairwise comparison revealed a significant effect between the SGwPRM and Immersive conditions.

There was no significant effect between the SGnoP and the other two conditions. Users felt that the robot was more of a collaborative partner in the SGwPRM condition.

The post experiment (PE) questionnaire was completed after all three conditions had been tested. Here, users ranked the three conditions in order of preference for the following questions.

- PEQ1: *I was aware of collisions as they happened*. There was a significant difference between conditions ($F_{2,27} = 12.47$, $p < 0.05$). Pairwise comparison revealed a significant effect between the SGwPRM and Immersive conditions, but no significant effect between the SGnoP and the other two conditions. Users felt that they were most aware of collisions while using the SGwPRM condition.
- PEQ2: *I had a feeling of working in a collaborative environment.* There was a significant difference between conditions ($F_{2,27} = 17.90$, $p < 0.05$). Pairwise comparison revealed a significant main effect between SGwPRM and the other two conditions, but no significant effect between the Immersive and SGnoP conditions. The SGwPRM condition was selected as providing the users with the greatest feeling of working in a collaborative environment.
- PEQ3: *I felt the robot was a partner*. There was a significant difference between conditions ($F_{2,27} = 17.90$, $p < 0.05$). Pairwise comparison revealed a significant main effect between SGwPRM and the other two conditions, but no significant effect between the Immersive and SGnoP conditions. The SGwPRM condition provided the users with a feeling that the robot was a partner.
- PEQ4: *The interface was intuitive to use.* There was no significant difference due to condition.
- PEQ5: *I was aware of the robot's surroundings.* There was a significant difference between conditions ($F_{2,27} = 8.39$, $p < 0.05$). Pairwise comparison showed a significant effect between the SGwPRM and Immersive conditions, but no significant effect between the SGnoP and the other two conditions. Users felt that the SGwPRM condition enabled them to be the most aware of the robot's surroundings.
- PEQ6: *I had to always pay attention to the robot's actions.* There was a significant difference between conditions ($F_{2,27} = 8.77$, $p < 0.05$). Pairwise comparison showed a significant effect between the Immersive condition and the two others, but no significant effect between the SGnoP and SGwPRM conditions. User felt that they needed to pay attention to the robot's actions most in the Immersive condition.
- PEQ7: *I felt the robot was a tool.* There was no significant difference between the three conditions.
- PEQ8: *I felt I was present in the robot's environment.* No significant difference was found between the three conditions.
- PEQ9: *I knew when the robot was about to collide with an object.* There was a significant difference between conditions ($F2,27 = 9.62$, $p < 0.05$). Pairwise comparison revealed a significant effect between the SGwPRM and the other two conditions, but no significant effect between the Immersive and SGnoP conditions. Participants felt that the SGwPRM condition was best for maintaining awareness of potential collisions.

V. Discussion

The Immersive condition was significantly faster than both the SGnoP and SGwPRM conditions. This result could be in part due to the lower learning curve of the Immersive condition. This hypothesis is supported by comments users provided in the post experiment questionnaire. Five users commented that the Immersive condition was simple and straight forward to use or that there was no learning curve.

In contrast, the SGnoP and SGwPRM conditions were a bit more difficult for the participants to become acquainted with. This higher learning curve is due to two issues. First, the user had to become familiar with the dialog that the system understood in a relatively short period of time. Second, at the same time the users also had to become familiar with selecting locations and objects in the AR environment.

Even though the users completed the task fastest in the Immersive condition, they also had the worst accuracy in this condition. Participants performed best in terms of accuracy in the SGwPRM condition. So although the SGwPRM condition took, on average, the longest time to complete the task, it resulted in the most accurate performance. It's not surprising to see that the SGwPRM has a longer completion time. This result is inherent in the design of the interface, as it takes time for the robot to display its plan in AR, for the user to agree with or modify the plan, and then have the robot execute the plan.

Although there was no significant effect of condition on the number of collisions, there was a significant effect on the number of close calls. The condition that performed the worst in this measure was the Immersive condition, while the SGwPRM condition performed the best. This result combined with the results from questions PTQ1, PEQ1, PEQ5 and PEQ9 indicate that the SGwPRM condition provided the users with the highest level of situational awareness.

An analysis of the dialog used revealed that deictic phrases, such as "go here", were used 87% of the time for the SGnoP condition and 93% of the time for SGwPRM. The remaining times deeper spatial dialog was used, such as "to the left of this" whilst selecting an object in the AR environment. This result of mainly using the deictic gestures could be due to the learning curve mentioned previously. To use the deeper spatial dialog the participants had to remember longer phrases and

coordinate issuing these phrases with the selection of objects in AR. Although this coordination is not difficult to master with practice, the participants tended to use a method that they could immediately master.

Another subjective measure was the feeling of working in a collaborative environment. The responses from questions PTQ6, PEQ2 and PEQ6 show that the users felt that they were working in a collaborative environment when completing the task using the SGwPRM condition. Question PEQ3 responses show that participants felt the robot was a partner when working with the SGwPRM condition. These results show that participants felt they were working in a collaborative team environment in the SGwPRM condition.

The last subjective question was to select the most effective condition. Nine of the participants selected the SGwPRM as the most effective, with one selecting SGnoP. Reasons provided for selecting SGwPRM included effective path creation, verbal feedback from the robot and the ability to change the plan mid-stream. Conversely, reasons given by the nine participants for not choosing the other two conditions included that the lack of planning caused crashes, that the Immersive condition lacked situational awareness and there was limited feedback from the robot.

VI. Conclusions

This paper presented an experiment conducted to evaluate the AR-HRC system. The experiment involved using three interfaces for working with a remotely located mobile robot. One interface was direct teleoperation where the user received visual cues from a camera mounted on the robot and drove the robot through direct teleoperation. A second interface provided the user with an exo-centric view of the robot in its work environment and enabled the human to use speech and gesture to communicate to the robot where it was to go.

The third interface provided the user with the same exo-centric view of the robot and allowed for spatial dialog and gesture interaction. However, this interface also enabled the human to collaborate with the robot to create, modify and review a plan before the robot executed it. This interface is the Augmented Reality Human-Robot Collaboration System.

Objective measures showed that the AR-HRC interface resulted in better accuracy and fewer close calls as opposed to the other two interfaces. The direct teleoperation interface resulted in the fastest time to completion, but did not fare as well as the other two interfaces for accuracy and close calls.

Subjective questioning showed that users felt they were working in a collaborative environment when using the AR-HRC interface. In this interface users also felt that they maintained better situational awareness, which is supported by the objective measurements of accuracy and close calls. Users also felt that the robot was more of a partner in the AR-HRC interface.

The users overwhelmingly selected the AR-HRC interface as the most effective of the three interfaces tested. The results of this study show that by providing the human with a shared view of the robots workspace and enabling the human to use natural speech and gesture, effective communication can take place between the robot and human. Common ground is easily reached by visually displaying the robots intentions in this shared workspace. Therefore, an environment has been created that allows for effective communication, and thus, collaboration.

References

[1] S. Thrun, "Toward a Framework for Human-Robot Interaction," *Human-Computer Interaction*, vol. 19, pp. 9-24, 2004.

[2] S. A. Green, X. Chen, M. Billinghurst, and J. G. Chase, "Collaborating with a Mobile Robot: An Augmented Reality Multimodal Interface," *17th International Federation of Automatic Control (IFAC-08) World Congress, July 6 - 11*, pp. Seoul, Korea, 2008.

[3] R. T. Azuma, "A Survey of Augmented Reality," *Presence: Teleoperators and Virtual Environments*, vol. 6, pp. 355-385, 1997.

[4] H. H. Clark and S. E. Brennan, "Grounding in Communication," in *Perspectives on Socially Shared Cognition*, L. Resnick, Levine J., Teasley, S., Ed. Washington D.C.: American Psychological Association, 1991, pp. 127 - 149.

[5] P. Milgram, S. Zhai, D. Drascic, and J. Grodski, "Applications of Augmented Reality for Human-Robot Communication," presented at In Proceedings of IROS 93: International Conference on Intelligent Robots and Systems, Yokohama, Japan, 1993.

[6] R. A. Bolt, "Put-That-There: Voice and Gesture at the Graphics Interface," *In Proceedings of the International Conference on Computer Graphics and Interactive Techniques*, vol. 14, pp. 262-270, 1980.

[7] M. Skubic, D. Perzanowski, S. Blisard, A. Schultz, W. Adams, M. Bugajska, and D. Brock, "Spatial language for human-robot dialogs," *Systems, Man and Cybernetics, Part C, IEEE Transactions on*, vol. 34, pp. 154-167, 2004.

[8] T. Fong, C. Thorpe, and C. Baur, "Multi-robot remote driving with collaborative control," *IEEE Transactions on Industrial Electronics*, vol. 50, pp. 699-704, 2003.

[9] C. Bowen, J. Maida, A. Montpool, and J. Pace, "Utilization of the Space Vision System as an Augmented Reality System for Mission Operations," *Proceedings of AIAA Habitation Conference*, pp. Houston TX, 2004.

[10] J. Maida, C. Bowen, and W. Pace, "Enhanced Lighting Techniques and Augmented Reality to Improve Human Task Performance," *NASA Tech Paper TP-2006-213724*, pp. July, 2006.

[11] J. Drury, J. Richer, N. Rackliffe, and M. Goodrich, "Comparing Situation Awareness for Two Unmanned Aerial Vehicle Human Interface Approaches," *Proceedings IEEE International Workshop on Safety, Security and Rescue Robotics (SSRR). Gainsburg, MD, USA* pp. August, 2006.

[12] S. A. Green, M. Billinghurst, X. Chen, and J. G. Chase, "Human-Robot Collaboration: A Literature Review and Augmented Reality Approach in Design," *International Journal of Advanced Robotic Systems*, vol. 5, pp. 1- 18, March, 2008.

[13] H. A. Yanco, J. L. Drury, and J. Scholtz, "Beyond usability evaluation: Analysis of human-robot interaction at a major robotics competition," *Human-Computer Interaction Human-Robot Interaction*, vol. 19, pp. 117-149, 2004.

[14] eMagin, *www.3dvisor.com* , last accessed June 2008.

[15] ARToolKit, *http://www.hitl.washington.edu/artoolkit/*, last accessed August 2008.

[16] A. G. Greenwald, "Within Subjects Designs: To Use or Not To Use?," *Pyschological Bulletin*, vol. 83, pp. 314 - 320, 1976.

[17] NIST and SEMATECH, "e-HandBook Engineering Statistics," http://www.itl.nist.gov/div898/handbook/prc/section4/prc47.htm, accessed August 2008.